
ALGORITMA K-MEDOIDS UNTUK PENENTUAN STRATEGI PEMASARAN PRODUK

Wiwit Agus Triyanto

Fakultas Teknik, Program Studi Sistem Informasi

Universitas Muria Kudus

Email: at.wiwit@yahoo.co.id

ABSTRAK

Strategi pemasaran produk merupakan hal yang sangat penting bagi perusahaan dimana strategi pemasaran merupakan pendekatan pokok yang digunakan oleh unit bisnis dalam mencapai sasaran yang di dalamnya tercantum keputusan-keputusan mengenai target pasar, penempatan produk, bauran pemasaran, serta tingkat biaya pemasaran yang diperlukan. Untuk dapat melakukan strategi pemasaran produk yang lebih efektif dan efisien, maka perlu dilakukan pengolahan data penjualan. Metode pengolahan data seperti ini sering disebut sebagai *data mining*. Pada penelitian ini akan menggunakan algoritma *K-Medoids* untuk pengelompokan data penjualan, sehingga akan ditemukan informasi yang dapat digunakan untuk penentuan strategi pemasaran produk yang tepat. Hasil dari penelitian ini menghasilkan 5 *cluster* dengan *cluster* pertama terdiri dari 909 *record* transaksi, *cluster* kedua terdiri dari 166 *record* transaksi, *cluster* ketiga terdiri dari 66 *record* transaksi, *cluster* keempat terdiri dari 132 *record* transaksi, *cluster* kelima terdiri dari 87 *record* transaksi. Strategi pemasaran produk dapat dilakukan dengan melakukan promosi pada *cluster* kelima yang memiliki kombinasi jumlah barang dibeli yang paling tinggi

Kata kunci: *strategi, pemasaran, produk, cluster, k-medoids*

ABSTRACT

Product marketing strategy is very important for companies where the marketing strategy is the basic approach used by the business unit in achieving the stated objectives in which decisions about the target market, product positioning, marketing mix, as well as the necessary level of marketing costs. To be able to do product marketing strategies more effective and efficient, it is necessary to processing of sales data. Data processing methods are often referred to as data mining. This research will use the K-Medoids algorithm for clustering of sales data, so it will be found that the information can be used to determine the right product marketing strategy. Results from this study resulted in five clusters with the first cluster consists of 909 records of transactions, the second cluster consists of 166 records of transactions, the third cluster consists of 66 record the transaction, the fourth cluster consists of 132 records of transactions, the fifth cluster consisting of 87 record the transaction. Product marketing strategy can be done with promotions on the fifth cluster that has a combination of the number of items purchased highest.

Keywords: *strategy, marketing, product, cluster, k-medoids*.

1. PENDAHULUAN

Strategi pemasaran produk merupakan hal yang sangat penting bagi perusahaan dimana strategi pemasaran merupakan suatu cara mencapai tujuan dari sebuah perusahaan. Strategi pemasaran adalah pendekatan pokok yang digunakan oleh unit bisnis di dalam mencapai sasaran yang di dalamnya tercantum keputusan-keputusan mengenai target pasar, penempatan produk, bauran pemasaran, serta tingkat biaya pemasaran yang diperlukan.

Untuk dapat melakukan pemasaran produk yang lebih efektif dan efisien, maka perlu dilakukan pengolahan data penjualan. Data yang telah didapatkan akan diolah untuk mengetahui pola dari data, sehingga dapat di ambil informasi yang tersembunyi dari data tersebut. Metode pengolahan data seperti ini sering disebut sebagai *data mining*. Pada penelitian ini analisa *data mining* dilakukan dengan menggunakan metode *clustering*.

Metode untuk melakukan *clustering* dapat dikategorikan menjadi empat metode, yaitu: *partitioning*, *hierarchical*, *grid-based* and *model-based*. *Clustering* berbasis *partitioning* menghasilkan partisi dari data

sehingga objek dalam *cluster* lebih mirip satu sama lain daripada objek yang ada dalam *cluster* lain. *K-Means* dan *K-Medoids* adalah contoh dari metode *partitioning*. Algoritma *K-Means* sensitif terhadap *outlier* karena objek dengan nilai yang sangat besar dapat secara substansial mendistorsi distribusi data. Untuk mengambil nilai rata-rata dari objek dalam sebuah *cluster* sebagai titik acuan, *medoid* dapat digunakan, yang merupakan objek dalam sebuah *cluster* yang paling terpusat. Strategi dasar dari algoritma *clustering K-Medoids* adalah untuk menemukan *k cluster* dalam *n* objek dengan pertama kali secara *arbitrarily* menemukan wakil dari objek (*medoid*) untuk tiap-tiap *cluster* [1].

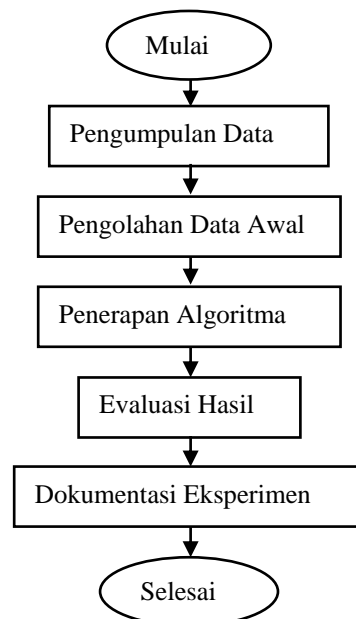
Pada penelitian ini akan menggunakan algoritma *K-Medoids* untuk pengelompokan data penjualan, sehingga akan ditemukan informasi yang dapat digunakan untuk penentuan strategi pemasaran produk yang tepat.

2. METODOLOGI

Metode penelitian dalam penelitian ini menggunakan metode eksperimen dengan tahapan seperti berikut:

- a. Pengumpulan data
- b. Pengolahan data awal
- c. Penerapan algoritma
- d. Evaluasi hasil
- e. Dokumentasi eksperimen

Adapun skema alur tahapan penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Skema Alur Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Teknik pengumpulan data dilakukan dengan mengambil *dataset* transaksi penjualan pada sebuah supermarket. Terdapat 304 atribut yang 303 merupakan *item*/barang dan salah satunya merupakan ID transaksi penjualan. Terdiri dari 1361 *record* transaksi. Adapun contoh data dari *dataset* dapat dilihat pada tabel 1.

Tabel 1. Contoh *Dataset*

Basket ID	Lemons	Standard coffee	Frozen Chicken Wings	98pct. Fat Free Hamburger	Sugar Cookies
C11867	false	false	true	true	false
C5096	false	false	false	true	false
C4295	false	false	false	false	false
C2837	true	false	false	false	true
C2693	true	false	true	false	false

Keterangan:

- Basket ID merupakan ID Transaksi penjualan barang.
- Lemons, Standard coffee, Frozen Chicken Wings, 98pct. Fat Free Hamburger, Sugar Cookies merupakan barang-barang yang dijual di Supermarket.
- False menandakan bahwa barang tersebut tidak dibeli oleh pembeli.
- True menandakan bahwa barang tersebut dibeli oleh pembeli.

3.2 Pengolahan Data Awal

Data yang sudah dikumpulkan akan diolah dengan beberapa tahap sehingga menjadi sebuah dataset yang akan diimplementasikan terhadap metode-metode *data mining*. Semua record transaksi yang ada pada dataset penjualan digunakan dalam tahap pengolahan data. Tahapan pengolahan data yang dilakukan yaitu:

- Data Preprocessing
Data *Preprocessing* yaitu dilakukan pembersihan dan persiapan data untuk menghilangkan konsistensi data, data tidak lengkap dan redundant data yang terdapat pada data awal. Data *Preprocessing* juga melakukan perubahan status barang yang semula *true / false* menjadi 1 / 0, yang akan digunakan untuk proses asosiasi.
- Menghitung Jumlah Barang yang Dibeli / Tidak Dibeli
Penghitungan jumlah barang yang dibeli maupun yang tidak dibeli dilakukan dengan menambah 2 atribut di dalam dataset, kedua atribut tersebut digunakan untuk melakukan proses *clustering* data.

Adapun hasil perhitungan jumlah beli.tidak beli ditunjukkan pada gambar 2.

Basket ID	Jumlah Barang yang dibeli	Jumlah Barang yang tak dibeli	Hair Conditioner	Lemons	Standard coffee
C11867	1	302	0	0	0
C5096	2	301	0	0	0
C4295	1	302	0	0	0
C2837	1	302	0	0	0
C2693	1	302	0	0	0
C3497	1	302	0	0	0
C2696	1	302	0	0	0
C1895	17	286	0	0	0
C9524	1	302	0	0	0

Gambar 2. Menghitung Jumlah Beli / Tidak beli

3.3 Penerapan Algoritma

Untuk melakukan *clustering* dengan metode partisi dapat menggunakan *K-Means* dan *K-Medoids*. *K-Means* merupakan suatu algoritma pengclusteran yang cukup sederhana yang mempartisi *dataset* kedalam beberapa *cluster* k [2].

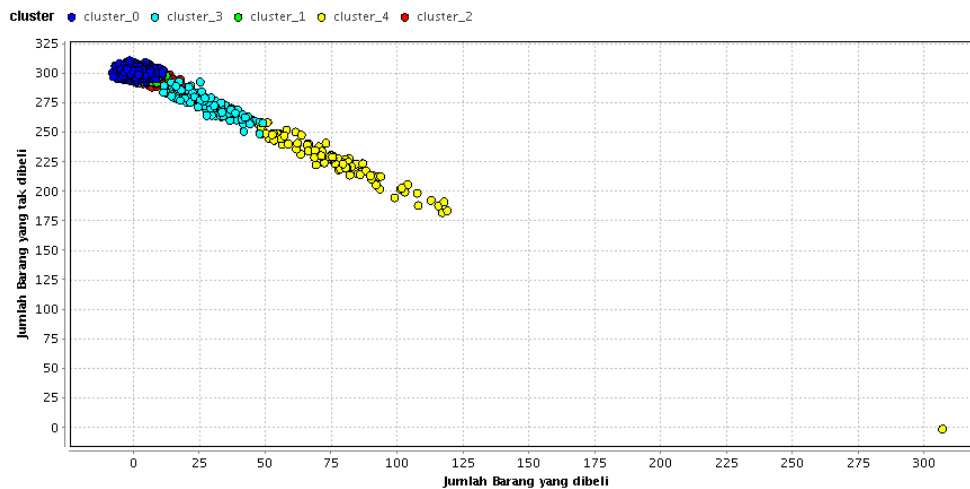
Algoritma *K-Medoids*, juga dikenal sebagai *partitioning around Medoids*, adalah varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan *Medoids* bukan dari pengamatan *mean* yang dimiliki oleh setiap *cluster*, dengan tujuan mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang ada dalam *dataset* [3].

Algoritma *K-Medoids* hadir untuk mengatasi kelemahan Algoritma *K-Means* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [4]. Menurut Han dan Kamber, algoritma *K-Medoids* adalah sebagai berikut [4]:

- Secara acak pilih k objek pada sekumpulan n objek sebagai *medoid*.
- Ulangi:
- Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid*.
- Secara acak pilih Orandom: sebuah objek *non-medoid*.
- Hitung total *cost*, S , dari pertukaran *medoid* oj dengan Orandom.
- Jika $S < 0$ maka tukar oj dengan Orandom untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- Hingga tidak ada perubahan.

3.4 Evaluasi Hasil

Dari hasil eksperimen *clustering* data penjualan, diperoleh grafik seperti yang ditunjukkan pada gambar 3.



Gambar 3. Grafik Hasil *Clustering* data Penjualan

Informasi yang bisa diperoleh dari hasil proses *clustering* tersebut adalah pada transaksi penjualan yang semula terdiri dari 1360 *record* transaksi, dapat dikelompokkan menjadi 5 *cluster*:

- Cluster* pertama terdiri dari 909 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 1 - 2; dan jumlah barang yang tidak dibeli sekitar 301 - 302.
- Cluster* kedua terdiri dari 166 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 3 - 7; dan jumlah barang yang tidak dibeli sekitar 296 - 300.
- Cluster* ketiga terdiri dari 66 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 8 - 14; dan jumlah barang yang tidak dibeli sekitar 289 - 295.
- Cluster* keempat terdiri dari 132 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 15 - 48; dan jumlah barang yang tidak dibeli sekitar 255 - 288.
- Cluster* kelima terdiri dari 87 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 49 - 303; dan jumlah barang yang tidak dibeli sekitar 0 - 254.

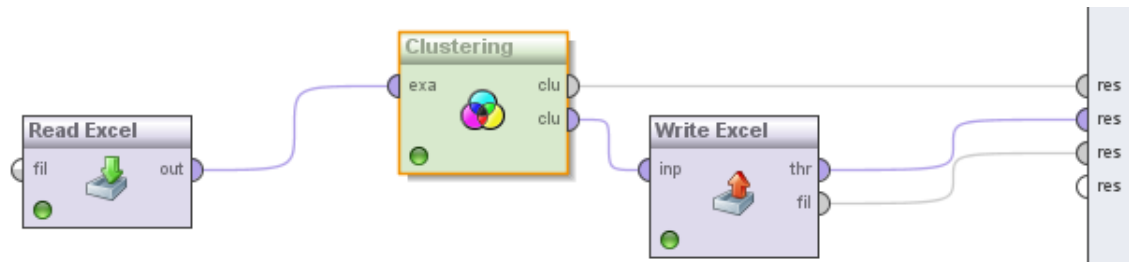
3.5 Dokumentasi Eksperimen

Dataset diambil dari data penjualan dalam bentuk *excel* diimport ke dalam *rapid miner*. Adapun hasil import data dari dataset ditunjukkan pada gambar 4.

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Basket ID	Jumlah Bar	Jumlah Bar	Hair Condit	Lemons	Standard cc	Frozen Chic	98pct. Fat F
polyno...	integer	integer	integer	integer	integer	integer	integer
attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute
C11867	1	302	0	0	0	0	0
C5096	2	301	0	0	0	0	0
C4295	1	302	0	0	0	0	0
C2837	1	302	0	0	0	0	0
C2693	1	302	0	0	0	0	0
C3497	1	302	0	0	0	0	0
C2696	1	302	0	0	0	0	0
C1895	17	286	0	0	0	0	1
C9524	1	302	0	0	0	0	0
C9527	1	302	0	0	0	0	0
C8726	2	301	0	0	0	0	0

Gambar 4. Import Data Dari Dataset

Kemudian data yang sudah diimport dilakukan proses *clustering* dengan membagi data tersebut menjadi 5 *cluster* (k). Hasil dari proses *clustering* tersebut akan di ekspor dalam bentuk file excel. Adapun proses *clustering* data ditunjukkan pada gambar 5.



Gambar 5. Proses Clustering Data

Adapun hasil *clustering* data ditunjukkan pada gambar 6.

Row No.	Basket ID	cluster	Jumlah Barang yang dibeli	Jumlah Barang yang tak dibeli
1	C11867	cluster_0	1	302
2	C5096	cluster_0	2	301
3	C4295	cluster_0	1	302
4	C2837	cluster_0	1	302
5	C2693	cluster_0	1	302
6	C3497	cluster_0	1	302
7	C2696	cluster_0	1	302
8	C1895	cluster_3	17	286
9	C9524	cluster_0	1	302
10	C9527	cluster_0	1	302
11	C8726	cluster_0	2	301
12	C9383	cluster_0	2	301
13	C8582	cluster_3	18	285
14	C10133	cluster_3	16	287
15	C6983	cluster_0	1	302
16	C3641	cluster_1	3	300
17	C5246	cluster_0	1	302
18	C10136	cluster_0	1	302
19	C7787	cluster_0	1	302

Gambar 6. Hasil Clustering Data

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan sebagai berikut:

- Pengelompokan menghasilkan 5 cluster dengan *Cluster* pertama terdiri dari 909 *record* transaksi, *Cluster* kedua terdiri dari 166 *record* transaksi, *Cluster* ketiga terdiri dari 66 *record* transaksi, *Cluster* keempat terdiri dari 132 *record* transaksi, *Cluster* kelima terdiri dari 87 *record* transaksi.
- Strategi pemasaran produk dapat dilakukan dengan melakukan promosi pada *cluster* kelima yang memiliki kombinasi jumlah barang dibeli yang paling tinggi.

DAFTAR PUSTAKA

- [1] T. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *International Journal Computer Technology & Applications*, vol. 3, 2012, pp. 1758-1764.
- [2] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. Mclachlan, A. Ng, B. Liu, P.S. Yu, Z.Z. Michael, S. David, and J.H. Dan, "Top 10 algorithms in *data mining*," *Knowledge and Information Systems*, 2008, pp. 1-37.
- [3] C. Vercellis, *Business Intelligence: Data mining and Optimization for Decision Making*, Milan: WILEY, 2009.
- [4] J. Han and M. Kamber, *Data mining: Concepts and Techniques*, Morgan Kaufmann, 2006.